



**CONFLICT: OPERATIONAL REALISM VERSUS ANALYTICAL RIGOR IN
DEFENSE MODELING AND SIMULATION**

GRADUATE RESEARCH PROJECT

Jason D. Horton, Major, USAF

AFIT/ISE/ENV/12-J04

**DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY**

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

DISTRIBUTION STATEMENT A. APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION IS UNLIMITED.

The views expressed in this thesis are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.

AFIT/ISE/ENV-12-J04

CONFLICT: OPERATIONAL REALISM VERSUS ANALYTICAL RIGOR IN
DEFENSE MODELING AND SIMULATION

Graduate Research Project

Presented to the Faculty

Department of Systems Engineering

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the
Degree of Master of Science in Systems Engineering

Jason D. Horton, B.S. Civil Engineering, M.B.A. Military Management

Major, USAF

June 2012

DISTRIBUTION STATEMENT A. APPROVED FOR PUBLIC RELEASE;
DISTRIBUTION IS UNLIMITED.

CONFLICT: OPERATIONAL REALISM VERSUS ANALYTICAL RIGOR IN
DEFENSE MODELING AND SIMULATION

Jason D. Horton, B.S. Civil Engineering, M.B.A. Military Management
Major, USAF

Approved:

Dr. Douglas D. Hodson (Co-Chairman)

Date

Dr. David R. Jacques (Co-Chairman)

Date

Abstract

This research seeks to identify sources of conflict between operational realism requirements and analytical rigor requirements in defense Modeling and Simulation (M&S) efforts, and to provide recommendations which help alleviate identified conflicts. This research focuses on methods that can be used to improve the development and implementation of operator in the loop (OITL) virtual environments intended for use in acquisition decision making or the evaluation of operational plans. It is believed that the reduction of conflict between operators and analysts will lead to a better use of scarce M&S resources and produce better analytic results from M&S studies used as a basis for defense acquisition decision making. A real-world defense acquisition M&S case study is provided as an illustrative example from which recommendations and lessons learned are derived.

Table of Contents

	Page
Abstract	iv
Table of Contents	1
Introduction	2
<i>How are operational realism requirements determined?</i>	2
<i>How does operational realism affect experimental outcomes?</i>	3
<i>What constitutes analytical rigor in defense experimentation?</i>	3
<i>Where does conflict between operational realism and analytical rigor exist?</i>	3
Background	4
Methodology	5
SIMAF Document Review	5
B-2 Pilot Data Gathering	6
Analyst Interview	6
Results - SIMAF Document Review Results	7
Technical Assessment Plan (TAP)	7
Analysis Traceability Matrix (ATM)	9
Requirements Traceability Matrix (RTM)	10
Operator Questionnaire and Interview Results	11
Analyst Interview Results	12
Analysis - Case Study Evaluated Against Selected GUIDEx Principles	15
Research Questions Answered	17
<i>How were the operational realism requirements determined?</i>	17
<i>How did operational realism affect experimental outcomes?</i>	18
<i>What constitutes analytical rigor with respect to this experiment?</i>	20
<i>What conflicts existed between operational realism and analytical rigor?</i>	22
Recommendations	27
Bibliography	28

Introduction

The purpose of this research effort is to identify sources of conflict between operational realism and analytical rigor in defense Modeling and Simulation (M&S) efforts, and to provide recommendations to help alleviate the identified conflicts. This research focuses on methods that can be used to improve the development and implementation of Operator In The Loop (OITL) virtual environments intended for use in acquisition decision making or the evaluation of operational plans. It is believed that the reduction of conflict between operators and analysts will lead to a better use of scarce M&S resources and produce better analytic results from M&S studies used as a basis for defense acquisition decision making. M&S in defense experimentation is an increasingly important tool used by the Department of Defense in making decisions concerning acquisition and force development. [1] In fact, it is DoD policy that “M&S is a key enabler of DoD activities.” [2] The importance of M&S to DoD decision making will continue to increase as budgets shrink, system complexity increases. This trend is especially apparent in the modeling of command and control systems, [3] which are well suited to computer modeling.

The four questions outlined below were central to this research effort. These questions were developed to help understand causes and effect of conflict between operational realism and analytical rigor.

How are operational realism requirements determined?

Modeling and simulation, by its very nature, is an abstraction of the physical world. Developers of OITL M&S environments must determine which portions of the

physical world they want to model, and with what degree of realism. As the resources available for M&S will always be finite, it is important that these resources be dedicated to developmental tasks that are central to answering the question posed by the customer or the hypothesis being tested, and not to areas that do not add value when determining the result of the simulation. M&S developers must also weigh competing requirements from multiple stakeholders when determining the level of realism that they will provide.

How does operational realism affect experimental outcomes?

The desired experimental outcome must be considered during the generation of the requirements for operational realism. Additionally, the level of realism must be considered when conducting the analysis of results. Did an unforeseen lack of realism affect the results in an un-predicted manner? Were the operators able to suspend belief sufficiently for their decision making to be evaluated?

What constitutes analytical rigor in defense experimentation?

In order for an experiment to be considered rigorous, and the results valid, the experiment should be designed using established criteria. First, we must determine what these criteria are. Once these criteria are known, we can identify places where these criteria are at odds with an operationally realistic M&S environment.

Where does conflict between operational realism and analytical rigor exist?

It is believed that the strict application of the experimental validity requirements mentioned above will often be in conflict with the desire to provide an operationally realistic environment. This paper seeks to identify these conflicts, and provide recommendation to help alleviate them.

Background

Different stakeholders will have different perspectives about what is important in a defense modeling and simulation (M&S) effort. In the author's experience, the operational community tends to place primary value on creating M&S environments which provide operational realism, while the analysis and modeling community places primary value on creating M&S environments that produce results which can be analyzed in a rigorous manner. Experience has shown that these objectives are often in conflict with each other.

The case study selected for this research effort was the B-2 Airborne Network Integration (ANI) Follow-On Analysis. This Modeling and Simulation (M&S) effort was conducted at Wright Patterson Air Force Base, Ohio, at the Simulation and Analysis Facility, (known as SIMAF), and the Global Strike Laboratory (GSL) at the Northrop Grumman Corporation (NGC) in El Segundo, California. The assessment was sponsored by the B-2 System Program Office (SPO). [3]

A primary reference for evaluation of the selected case study was the Guide for Understanding and Implementing Defense Experimentation, known as the GUIDEx. [1] This volume was produced by The Technical Cooperation Program (TTCP), which consists of representatives of defense science and technology (S&T) organizations in Australia, Canada, the United Kingdom, and the United States. Information contained in the GUIDEx used in evaluation of the selected case included experimental validity requirements, threats to a good experiment, and principles which should be considered in the development of a good experiment.

Methodology

This research was conducted using a literature review and case study format. The literature review was used primarily in the development and refinement of the research questions presented above. This review helped to establish the baseline for what constitutes an analytically rigorous experiment, and the importance of operational realism in determining experimental outcomes. The case study is employed as an illustrative example, in order to generate real world examples of conflict between operational realism and analytical rigor. After observing examples of these conflicts in a real world application, the recommendations for conflict mitigation found at the end of this article were developed.

Due to classification requirements, the details of the experiment and the results of the individual runs conducted during the course of the experiment will not be presented in this paper. This paper focuses on operator and analyst lessons learned at the unclassified level. Although the observed experiment was a distributed event, this research focused primarily on the portion of the case study conducted at the SIMAF facility at Wright-Patterson AFB due to funding, geographic, and access limitations.

SIMAF Document Review

The initial research stage was to review the documents provided by SIMAF, with the goal of determining what the operational realism requirements were, and how they were developed. Several SIMAF provided documents were reviewed in order to understand the objectives and method employed in this experiment. The primary document provided for review was the *B-2 Airborne Network Integration Follow-on Analysis Technical Assessment Plan*, referred to as the TAP. This document provides an

overview of the M&S effort, objectives for the assessment, and the analysis plan. In addition, the Requirements Traceability Matrix (RTM) and Analysis Traceability Matrix (ATM) were reviewed. [3]

B-2 Pilot Data Gathering

There were three B-2 pilots who participated in the experiment at the SIMAF facility. These pilots were the operators of the SIMAF B-2 for the exercise, and were interviewed by the author at the conclusion of the week-long event. The goal of this interview was to determine what the pilots thought the required level of operational realism was to meet the objectives of the M&S effort, and to what degree these requirements were met. In addition to the interview, the pilots were administered a written survey, designed to capture their reactions regarding the level of realism present in the simulation. The survey instrument contained a mixture of objective ratings and subjective evaluations of the hardware, software, and mission scenarios.

Analyst Interview

The author also interviewed the lead analyst for the exercise, and software integration lead for the effort. The goal of these interviews was to determine what the operational realism requirements were from the analyst's perspective, and to note the differences compared to the results obtained from the pilots. In addition, the analysts were asked to describe who they felt primary stakeholders were, and what impact they felt the operator experience and proficiency had on experimental outcomes.

Results - SIMAF Document Review Results

Technical Assessment Plan (TAP)

The TAP is the primary document which describes the M&S activities associated with this case study. The TAP defines the purpose of the activity studied as the evaluation of the “ integration of advanced tactical data link capabilities with United States Air Force (USAF) advanced platforms (designated as 5th generation) to assess the mission effectiveness realized when all aircraft are able to communicate during Anti-Access/Area Denial (A2AD) operations.” [3] Specifically, the event looked at the potential impact of improved communications capability on B-2 mission effectiveness.

As part of the overall effort, a series of constructive and interactive studies were conducted. These documents were primarily used for development of the simulation environment, and to help focus the main activity, which was a distributed virtual event. This event involved current and qualified B-2 pilots at both the SIMAF facility at Wright-Patterson AFB, and the NGC Global Strike Laboratory facility in El Segundo California. [3] The activity surrounding preparation and execution of the simulation environment and the B-2 “crew-cab” or cockpit at the SIMAF is the focus of this case study.

The primary purpose of the experiment was to “provide a mission environment, guided by B-2 Operator input” to assess the contribution of improved communications capability on mission effectiveness. There were three versions of B-2 simulation present: the SIMAF B-2, the GSL B-2, and Desktop B-2. The Desktop B-2 performed a supporting role, and was not used for mission effects assessment, due to its limited functionality. [3] The SIMAF B-2 is a B-2 cockpit simulator that includes actual aircraft

hardware as the user interface, and uses re-hosted software developed at the SIMAF facility for this exercise.

The TAP contains the objectives of the study, the analysis plan to include the measures employed, and a description of the mission vignettes that were employed during the virtual and constructive phases of the overall study. The overarching objective of the study was to assess how B-2 survivability, lethality, and situational awareness are affected by the addition of improved communications capability. Additionally, the study sought to assess the effect of this new capability on campaign-level objectives, such as the number of sorties and operations tempo required to prosecute a given target set.

The TAP also contains the assumptions and constraints [3] that apply to the analysis plan. The three that are most applicable to the research questions in this paper are:

- Every trial will run in real-time due to the fact that human-in-the-loop effects will be measured in this assessment.
- Environmental factors, such as route and Red air locations should be varied to prevent aircrew learning. These factors need to be included in the experimental design to account for the effect that varying them has on results.
- Due to the short time available for the [Virtual Event], only a small number of runs of each factor level and each case will be possible. Therefore, in-depth statistical analysis will be performed only during the High Side Studies.

The TAP includes the experimental design matrix developed for the event. The design matrix consisted of three levels of communication capability, three operational vignettes, and three crewmember configurations. Since there are 3 factors being evaluated, each with 3 levels, the full factorial [5] case matrix consists of 3^3 , or 27 cases.

The TAP analysis plan specifies that the cases be executed in a random order to minimize the effect of aircrew “gaming the system” from one case to the next. [3] “Gaming” in this instance refers to the ability of the operator to anticipate what is going to happen based upon previously executed trials.

Analysis Traceability Matrix (ATM)

The ATM description is found in the appendix of the TAP. It lists the analysis questions, hypotheses, and methodology used to generate the measures of performance associated with these analysis questions. The main analysis question is “How much does the integration of Line of Sight (LOS) and Beyond Line of Sight (BLOS) communication links on the B-2 result in better mission effectiveness during operations in AD [Area Denial] airspace?” The ATM describes the measures used to grade the objectives described above. [3]

Requirements Traceability Matrix (RTM)

The RTM traces requirements from the analysis questions found in the TAP to the weapons system and its associated functional requirements. The RTM breaks the weapons system capability being designed or modeled into eleven categories. Table 1 below provides a summary of the RTM. The far right column shows the number of requirements in each category.

Table 1: Requirements Traceability Matrix (RTM) Summary

Category	Name	Requirement Description	# of Requirements
1.0	Command and Control (C2)	Execute off-board commands and automatic modes for subsystem control	7
2.0	Fly and Maneuver (FM)	Aerodynamically steer the aircraft for preplanned and reactive routes or maneuvers.	4
3.0	Communicate (CM)	Send and receive messages via voice or digital data link	20
4.0	Understand, Predict, React (UP)	Present data to the operator. Assess and decide upon appropriate aircraft or system action	60
5.0	Sense and Detect (SD)	Sense via RF, IR, visual sensors; Assimilate data into detections and tracks using data fusion; pass info to the operator	6
6.0	Special category (SC)	Everything not covered by categories 1-5 and 7-11.	0
7.0	Launch Munitions (LM):	Release and support munitions. What's in LM? All weapon data.	0
8.0	Electronic Warfare (EW)	Employ electronic (RF) devices	0
9.0	Directed Energy Attack (DE)	Employ energy / optically based devices	0
10.0	Infrared Attack & Support (IR)	Employ IR based attack devices (DIRCM)	0
11.0	Instrumentation (IN):	Instrument the model, as required	6

Operator Questionnaire and Interview Results

The questionnaire administered to the operators yielded several key insights. First, it was noted that the three pilots interviewed had different opinions of the realism provided by the B-2 simulator at SIMAF. For example, two of the pilots rated the “hand flying” quality of the SIMAF B-2 simulator compared to the aircraft at 7 out of 10, while the third pilot gave it only 3. There was also a disparate view of how representative the weapons delivery procedures were, with two pilots rating them a 7 out of 10, and the third pilot giving them only 2 out of 10. Table 2 below summarizes the realism scores given by the B-2 operators at the conclusion of the experimental trials.

Table 2: B-2 Pilot Simulator Realism Questionnaire Responses

Question	1-10, where 10 is more realistic		
	Pilot A	Pilot B	Pilot C
The simulator “hand flies” like the aircraft.	7	7	3
The displays in the simulator are representative of the aircraft.	7	5	7
Weapons delivery procedures in the simulator are representative of the aircraft.	7	2	7
The threat and mission scenario presented is a realistic B-2 mission scenario.	6	7	8
The behavior and modeling of the support assets in this experiment was realistic.	6	4	9

Short answer questionnaire and interview responses from Pilot A revealed that he felt that simulator was representative enough to “validate the exercise.” He felt that the level of task saturation found in the scenario was low compared to what he would expect in a real scenario, largely because the vignettes that were utilized in the simulation represented only a small portion of a typical B-2 mission. He felt strongly that lessons

learned from one scenario had an effect on operator performance in subsequent scenarios, and noted that the event became “more of a conditioning experience / exercise by the end of the week.”

Pilot B felt that one of the biggest shortfalls in the B-2 simulator was the ability to deliver weapons in a realistic manner. He felt that the bare minimum weapons delivery functionality was present. He noted that crew errors and procedures could not be effectively evaluated because of the limited functionality. He pointed out that dynamic and time sensitive targeting procedures could not be executed given the available functionality. He felt that it was necessary to have an experienced operator involved in the simulation in order to “work around limitations and understand the real potential for this capability.”

Pilot C felt that many of the displays in the simulator were similar to actual aircraft displays, but lacked accuracy. He noted several issues related to weapons delivery procedures and displays that were not operationally realistic. He thought that some of the imposed scenario restrictions, such as the inability to communicate with outside agencies without the use of the new communications capability being evaluated, were unrealistic. He thought that pilot proficiency and decision making ability played a key role in the outcome of the experiment.

Analyst Interview Results

The interview session conducted with the analysts yielded several key insights. In general, they felt that most of the operational realism requirements which were actually specified were met. They noted, however, that the pilots expressed a desire for more

realism in several areas during the out brief. The also noted that the pilots each had a different view of which operational realism requirements were most important, and where the biggest shortfalls in operational realism were found in the SIMAF B-2 simulator.

A key takeaway from discussion with the analysts was that many of the realism shortfalls identified by the pilot were not captured in the requirements documents. In other words, the primary complaints that the pilots had were not due to un-met requirements; rather, they were due to requirements that were never identified in the first place. The analysts also identified differences of opinion between pilots as one of the challenges present when determining the requirements for operational realism. Different pilots will have different experience levels and operational backgrounds, which will influence what they consider acceptable from a realism perspective.

The analysts were asked what effect pilot experience and proficiency had on the outcome of this experiment. They noted that in many cases, the pilots found ways to accomplish the mission that were not anticipated prior to the event. An example they gave was how the pilots handled the emergence of certain “pop-up” or un-planned surface-to-air threats along the planned route of flight. The analysts and experiment designers had anticipated a number of reactions that the pilots might have when presented with this stimulus, and had planned to evaluate the effect that the capability being tested had on the efficacy of those reactions. During the event, the pilots performed a set of reactions based on the pop-up threat that were not anticipated in advance by the analysts.

The analysts also noted some instances where un-required realism was implemented. For example, early in the development of the SIMAF B-2, there was some effort placed towards the implementation of a functional landing gear system, even

though the simulation for the event was planned to take place entirely in the airborne environment, with no takeoffs or landings needing to occur. Similarly, operational fuel management displays and algorithms were implemented, even though the planned mission vignettes were short enough to make fuel management an inconsequential factor to mission success.

The analysts were asked “how were the requirements for operational realism determined?” They felt that early in the development process, there was little B-2 pilot involvement. The requirements for simulator functionality were largely determined by a team of contractors in the software integration and analysis roles. The analysts felt that more operator involvement, especially early in the process, would help avoid “going off on tangents” and using development time and resources for items which are not value-added with respect to satisfying the objectives and providing operational realism.

Analysis - Case Study Evaluated Against Selected GUIDEx Principles

The GUIDEx contains 14 principles for effective defense experimentation. [1] In order to help determine the validity of the experiment considered in this study, five of the principles were selected to evaluate the case study against. A summary of the selected principles is shown in Table 3 below.

Table 3: Selected GUIDEx Principles

	Principle Summary / Thesis
Principle #1	Defense experiments are uniquely suited to investigate the cause-and-effect relationships underlying capability development.
Principle #2	Designing effective experiments requires an understanding of the logic of experimentation.
Principle #7	Multiple methods are necessary within a campaign in order to accumulate validity.
Principle #8	Human variability in defense experimentation requires additional experiment design considerations.
Principle #14	Frequent communication with stakeholders is critical to successful experimentation.

Principle #1 is concerned with cause-and-effect relationships. In this case study, the primary cause under evaluation was the three different levels of communication capability employed. The TAP states that the main hypothesis for this case study is that “using [this capability], mission effectiveness, lethality, and survivability should be substantially increased.” [3]

Principle #2 is concerned with the development of effective experiments. In the discussion of the second principle, the GUIDEx lists “21 Threats to a Good Defense Experiment.” Failure to avoid these threats during development will lead to an experiment which is less effective. The threats that the author determined to be most applicable to the effects of operational realism on experimental outcomes are shown in Table 4 below.

Table 4: Threats to a Good Experiment

GUIDEx #	Threat	Question	Study Results
1	Capability not workable	Do the hardware and software work?	Capability under evaluation was sufficient. Weapons delivery functionality was lacking.
2	Player non-use	Do the players have the training and TTP to use the new capability?	2 of 3 B-2 pilots involved had no prior experience using new communications capability prior to VE #3.
18	Non-representative capability	Is the experimental surrogate functionally representative?	Unknown; new capability has not yet been implemented. Real world performance may not match experimental.
19	Non-representative players	Is the player unit similar to the intended operational unit?	Current, qualified operators were used, but experience level was above average compared to operational units.
21	Non-representative scenario	Are the Blue, Green, and Red conditions realistic?	Scenario was considered restrictive by the pilots, with limited threats.

Principle #7 states that multiple methods should be used to accumulate validity, because there is “no such thing as a perfect experiment.” [1] This case study used a combination of constructive simulations and OITL simulations in the overall analysis effort. The constructive simulations were used primarily to aid the selection of the three factors that were used in the experimental design, and also to help focus the vignettes that were used in the event.

Principle #8 is concerned with human variability. In this case study, the sample size of three B-2 pilots is extremely small. As the GUIDEx notes, “because humans are unique, highly variable and adaptable in their response to an experimental challenge, they are more than likely to introduce large experimental variability.” [1] The potential for variability is exacerbated by the small sample size in this case.

Principle #14 highlights the “importance of engaging in continuous dialogue with stakeholders.” [1] This research identified that one of the hindrances to providing an appropriate level of operational realism was a lack of effective early communication with stakeholders from the operational community.

Research Questions Answered

Now that the results of this research effort have been presented, we will return to the four research questions that were asked in the introduction to this paper. These questions will be answered in terms of the case study presented above.

How were the operational realism requirements determined?

The method used to determine the operational realism requirements for a given M&S effort will have a great effect on what those operational realism requirements are, and in turn on the outcome of the M&S effort as a whole. It is important to identify which of the many stakeholders involved have inputs into operational realism requirements, and to make sure that those inputs are properly captured. It is important to note that the most of the stakeholders who care about the level of operational realism, such as the operators themselves, probably have little knowledge of what it takes to develop an M&S environment. Software developers and weapon systems operators

typically speak different languages. This makes the process for determining where realism is required (and where it's not) a crucial factor in the eventual success or failure of an M&S effort.

The operational realism requirements for this case study were largely determined by a team of contractors. At the beginning of the program, there was little involvement by operational B-2 pilots in determining the operational realism requirements. The developers generally felt that communication was lacking in the early stages of the project.

One of the effects of this lack of early involvement can be seen in Table 1 above, which summarizes the Requirements Traceability Matrix (RTM). As can be seen, the primary emphasis, as one would expect for an M&S effort of this type, was on requirements necessary for the implementation of the improved communications capability in the simulation. This is exemplified by the fact that there are 20 requirements related to the Communicate (CM) capability area, and 60 requirements associated with the Understand, Predict, React (UP) area. In comparison, there are zero requirements associated with the Launch Munitions (LM) capability area.

How did operational realism affect experimental outcomes?

In some OITL M&S scenarios, the most important results may be the subjective observations of the operator regarding the new capability or tactic being employed. This is particularly true when the operator's situational awareness or knowledge of the threat environment is being evaluated. When subjective operator feedback is data used to determine the experimental outcome, the level of realism present is likely to have a large effect. In general, operators lend more credence to scenarios with realism than they

would to unrealistic scenarios or environments. In this type of scenario, the developers may find that they need to satisfy the operator's realism requirements in order for the outcomes of the experiment to be considered valid by the operational community.

In M&S environments where the key results are objective measurements, the level of operational realism present will still affect the experimental outcome. For example, if the number and type of threats found in a scenario is not representative, the survivability and mission effectiveness of the system under evaluation cannot be effectively determined. Realism is particularly important when evaluating an operator's use of a new capability. As an example, consider an M&S environment established with the objective of determining whether the user interface of a new targeting pod is suitable for use by the pilot of a single-seat aircraft. If the environment and mission tasks presented to the operator are not realistic, the experimental outcome may not be representative. In order to achieve a valid result from an experiment such as this one, the operator would need to be able to perform an operationally realistic number and type of tasks during the mission scenario, such as flying the aircraft, operating the radar, and transmitting and receiving radio communications. Operational realism is particularly important to achieving valid experimental outcomes when the experiment has the operator-in-the-loop (OITL) and the outcome of the experiment depends on the ability of the operator to process information and make correct decisions based on the information presented to him.

Data collected from the operator surveys and interviews revealed that weapons delivery capability was one area where the B-2 pilots felt that a lack of operational realism hindered the ability to effectively evaluate the capability being studied. Pilot B

noted that the provided functionality did not allow for realistic Dynamic Targeting (DT) or Time Sensitive Targeting (TST) procedures in the B-2 simulator. The effect of varying levels of communications capability on DT and TST was a key part of the measures used to evaluate the objectives.

The pilots also had varying opinions when asked “did a lack of fidelity in the B-2 simulator detract from your ability to effectively evaluate the capability being studied in this exercise?” Pilot A felt that it had no effect, and listed no faults with the simulator hardware or software that he felt detracted from the exercise. Pilot B answered “Yes” to the lack of fidelity question. He pointed out several limitations in areas such as flight plan management, stores management, and threat situation displays that he felt made capability evaluation more difficult. Pilot C felt that “overall, the simulator served its purpose,” but felt that the inability to edit the planned routing and weapons settings, along with the aircraft communications systems, detracted from the realism of the simulator. As can be seen, the different experience levels and expectations of each operator will affect the level of realism required for that operator to feel that the results of the experiment can be safely generalized to real world operations beyond the confines of the experiment itself.

What constitutes analytical rigor with respect to this experiment?

All experiments consist of the components shown in the Table 5 below. [4] These components will be discussed as elements of the case study presented below.

Table 5: Components of an Experiment

Component	Definition
Treatment	The possible cause; a capability or condition that may influence warfighting effectiveness.
Effect	The result of the trial, a potential increase or decrease in some measure of warfighting effectiveness.
Experimental unit	Executes the possible cause and produces an effect.
Trial	One observation of the experimental unit under treatment (or lack of treatment) to see if the effect occurred.
Analysis	Compares the results of one trial to those of another.

A good experiment contains the information necessary to determine whether or not the applied treatment caused an effect. Table 6 below reflects four logically sequenced requirements necessary to achieve a valid experiment for an OITL-centric defense M&S effort. [1]

Table 6: Experiment Validity Requirements

Ability to...	Discussion
Use the new capability.	For the experiment to be valid, the operator must be able to employ the new capability under relevant conditions. Hardware and software must work as anticipated.
Detect a change in the effect.	Experimental error may produce too much variability in results. Reduction of experiment variation through limited stimuli presentations and a controlled external environment mitigate experiment-induced error.
Isolate the reason for a change in the effect.	Was the observed change in effect due to the intended cause (i.e., the new capability) or something else? Are the results confounded by some alternative explanation?
Relate the results to actual operations	The ability to apply results beyond the experimental context pertains to the experiment realism and robustness. Are the results applicable to operational forces in actual military operations?

In this experiment, analytical rigor was established by adhering to the four experimental validity requirements. The operator(s) were able to employ the new capability under relevant conditions. Measures were selected such that a change in effect could be detected at various treatment levels. The ability to relate the results to actual operations was provided through the use of a realistic mission scenario, and the use of current and qualified weapon system operators in the experiment. This experiment also adhered to an accepted factorial design [7] process.

What conflicts existed between operational realism and analytical rigor?

The application of the experimental validity requirements above leads to several potential conflicts between operational and analytical stakeholders. The first requirement is that the new capability must be employed “under relevant conditions.” It is likely that the operational and analytical communities will be at odds when defining what these relevant conditions are. In general, the operator is likely to consider a wide range of conditions to be relevant. For example, consider the case where we are trying to assess the effectiveness of a new targeting pod for a fighter aircraft. The operator stakeholder is likely to consider a wide range of conditions to be relevant, such as day and night, high and low altitude, high humidity and low humidity, etc. The operator may feel that the inclusion of all of these conditions is necessary to achieve a valid result, without fully appreciating that doing so will make the developers and analysts jobs much more difficult by increasing the number of variables which must be considered in order to detect and isolate the reason for the change in effect produced. Increasing the number of conditions will also invariably increase the cost and development time of the scenario, neither of which is unlimited. Therefore, it is important to determine early on in the development

process which conditions and variables are most important operationally, and to develop the M&S environment to support these conditions.

The requirement to be able to detect a change in effect may also lead to conflict. The discussion in table 6 above notes that it may be necessary to employ techniques such as the “reduction of experiment variation through limited stimuli presentations” and to “provide a controlled external environment to mitigate experiment-induced error.” The exclusion of stimuli and the provision of a controlled environment are two things that are not found in an operationally realistic combat scenario. Therefore, the operator’s desire for operational realism (i.e., a dynamic and uncertain environment) may increase experimental noise and make it difficult for the analyst to measure a change in the effect produced by the new capability.

The third requirement, that we have the “ability to isolate the reason for a change in the effect” may also lead to conflict. The inclusion of actual operators in the experimental design tends to increase the realism and applicability of the experiment, but can also lead to other problems. First, if the operators are knowledgeable of what capability is being measured, they may bring their own biases concerning the capability to the experiment, which can affect their performance during the simulation both positively and negatively. Also, using actual operators may lead to confounding results from issues such as the “learning effect” [1] between subsequent trials, which may make it difficult to isolate the reason for a change in the effect.

The fourth experimental validity requirement, that we have the “ability to relate the results to actual operations” will lead to conflict between operators and analysts. Generalizing the results of the experiment beyond the context of the experiment itself

requires the “representation of surrogate systems, the use of operational forces as the experimental unit, and the use of operational scenarios with a realistic reactive threat.”

[1] Each of these things is desirable to the operator because they increase operational realism and applicability of the scenario. The actual future combat scenario encountered by the system being modeled is unlikely to be exactly like the modeled scenario.

Increased realism allows the M&S scenario results to be generalized further beyond the context of the experiment itself. The price of this increased realism is to make the analyst’s job of detecting and isolating the reason for the change in effect more difficult, because of the increase in the number of variables present.

The “use of operational forces as the experimental unit” also leads to conflict. The use of current and qualified systems operators in OITL M&S environments will increase the realism of the experiment because it helps to ensure that realistic tactics and procedures are employed, and that the level of operator proficiency employed in the simulation is commensurate with that found in current fielded forces. However, the use of current operators introduces another variable which must be accounted for. This is because a range of operator proficiency levels will tend to produce a range of experimental results. Particularly in experiments which involve a small sample size of operators, it may be difficult for the analyst to determine the impact that operator proficiency has on the experimental outcome, which can lead to confounded results.

The necessity of using “operational scenarios with a realistic reactive threat” can lead to an area of conflict between operational realism and analytical rigor. A threat which is able to vary its response based on the action of the blue capability under study is more realistic than a threat which performs the same way in every scenario, or one which

makes its inputs in a scripted manner. However, the operator's desire for a "thinking" threat introduces another source of variability, which makes the analyst's job of detecting and isolating the reason for a change in the effect more difficult.

In an article discussing conceptual modeling, Robinson [8] notes that modelers are primarily concerned with the concept of *validity*, which he defines as "a perception, on behalf of the modeler, that the conceptual model can be developed into a computer model that is sufficiently accurate for the purpose at hand." [8] This desire for validity often leads the modeler, and by extension the analyst, to prefer a tightly controlled environment. Conversely, he notes that the client, who is often the operator in defense M&S environments, is primarily concerned with *credibility*. He defines model credibility in a similar manner to validity, with the key distinction that the credibility of the model depends on the client perception, versus the modeler's perception. In order for the model to be credible, the client must be convinced that "all the important components and relationships are in the model." [8]

The selected case study highlights several areas of potential conflict between operational realism and analytical rigor. First, the number of "runs" or scenarios that could be accomplished during the experiment was limited by the operators. In this case, the pilots were only available for one week of activity, due to manning constraints and the fact that they needed to travel to Wright Patterson Air Force Base from their home station. The length of the planned mission scenarios limited the number of scenarios that could be reasonably be accomplished in a day to about six. This limited the number of runs that could be accomplished in a five day exercise week to about 30. Discussion with

the analysts revealed that the selection of a $3^3 = 27$ factorial experimental design matrix was made with these limitations in mind.

As can be seen, the necessity of using actual operators to provide realism can be limiting from the analysts perspective. In this case, the number of factors selected for evaluation was limited to three factors with three levels each. Given these constraints, most of the cases in the design matrix were only able to be run once. This makes the elimination of confounding factors more difficult, since each run has a different set of factors. The inability to run each case multiple times causes the results of each case have less weight.

Scenario limitations were another source of conflict between operators and analysts. As mentioned, the operators found ways to accomplish the mission that were not anticipated by the analysts. The ability for the operator to maneuver in a manner that he considers to be the most tactically sound provides an increase in operational realism. However, this freedom of maneuver also causes difficulties for the analyst. In this case, the analysts were not able to evaluate effectiveness of the anticipated threat reactions, because the pilots didn't take the anticipated actions. Instead, the analysts were forced to evaluate the effectiveness of the actions that the pilots actually took.

The requirement for realism in "supporting systems" is also another source of conflict. As one would expect, the developers of this M&S environment were primarily concerned with ensuring that that capability being evaluated worked as required. As highlighted earlier, this can be seen in the RTM, where nearly all the requirements listed are tied to the new capability being tested, and almost none of the requirements are associated with ensuring that the B-2 simulator can perform existing, basic aircraft

functions. In some cases, the lack of basic B-2 simulator functionality reduced the overall operational realism of the scenario.

Recommendations

The evaluation of the case studied for this article led to several recommendations which may be used when designing OITL M&S environments to help reduce the conflict between operational realism and analytical rigor.

1. When planning an OITL experiment, identify key operator stakeholders early, and establish a communication plan with them.
2. Develop operational realism requirements for each mission objective.
3. Get operators to interact with actual hardware and software at key intervals during the development process to ensure that realism requirements are being met.
4. Understand that operators will assume basic functionality in any simulation and expect everything to work. Ask pointed questions to ascertain which weapon system functions are required to evaluate the new capability.
5. Eliminate unnecessary functionality (with operator concurrence) early in the development process to focus effort on functionality that aids in satisfying the objectives.

Bibliography

- [1] The Technical Cooperation Program (TTCP), Guide for Understanding and Implementing Defense Experimentation (GUIDEx), Ottawa, 2006.
- [2] United States, *DoD Modeling and Simulation (M&S) Management: DoD Directive 5000.59*, Washington D.C.: Department of Defense, 2007.
- [3] D. S. Alberts, J. J. Garstka, R. E. Hayes and D. A. Signori, Understanding Information Age Warfare, DoD Command and Control Research Program (CCRP), 2002.
- [4] Aeronautical Systems Center (Simulation and Analysis Facility), *B-2 Airborne Network Integration (ANI) Follow-On Analysis Technical Assessment Plan*, Wright-Patterson Air Force Base: United States Air Force, 2011.
- [5] L. S. Cortes, W. Duff and D. Bergstrom, "A primer on Design of Experiments for US Navy T&E Managers," Colorado State University, Fort Collins, 2011.
- [6] W. R. Shadish, T. D. Cook and D. T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*, Boston: Houghton Mifflin Company, 2002.
- [7] R. T. Johnson, G. T. Hutto, J. R. Simpson and D. C. Montgomery, "Designed Experiments for the Defense Community," *Quality Engineering*, vol. 24, pp. 60-79, 2011.
- [8] S. Robinson, "Conceptual Modeling for Simulation Part I: Definition and Requirements," *Journal of the Operational Research Society*, vol. 59, pp. 278-290, 2008.
- [9] D. J. Cloud and L. B. Rainey, *Applied Modeling and Simulation: An Integrated Approach to Development and Operation*, New York: McGraw-Hill, 1998.

Vita

Jason D. Horton

Contact Information

AFIT IDE Student
Air Force Institute of Technology
2950 Hobson Way, Bldg. 641
Wright-Patterson AFB, Ohio 45433
Phone (work): 937-255-6565 x4596
Phone (mobile): 660-441-1704
Email: jason.horton@afit.edu

Education

- 2012 **Master Of Systems Engineering (Residence)**
Air Force Institute of Technology, Wright-Patterson AFB, OH
Concentration: Airborne Systems
- 2009 **Master of Business Administration (MBA)**
TUI University International, Cypress, CA
Concentration: Military Management
- 2008 **Graduate, United States Air Force Weapons School**
Nellis Air Force Base, NV
- 2008 **Air Command and Staff College (Correspondence)**
Maxwell Air Force Base, AL
- 2005 **Squadron Officer School (Residence)**
Maxwell Air Force Base, AL
- 1999 **Undergraduate Pilot Training**
Columbus Air Force Base, MS
- 1998 **Bachelor of Science (B.S.) in Civil Engineering**
United States Air Force Academy, Colorado Springs, CO
Concentration: Structural Engineering

Professional History

- 1998 – **Air Force Officer and Pilot, B-2 Weapons Instructor Pilot, B-52 Pilot**

present

REPORT DOCUMENTATION PAGE*Form Approved*
OMB No. 074-0188

The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) 14 Jun 2012			2. REPORT TYPE Graduate Research Project		3. DATES COVERED (From – To) 19 May 2011 – 14 Jun 2012	
4. TITLE AND SUBTITLE Conflict: Operational Realism versus Analytical Rigor in Defense Modeling and Simulation					5a. CONTRACT NUMBER	
					5b. GRANT NUMBER	
					5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Horton, Jason D, Major, USAF					5d. PROJECT NUMBER N/A	
					5e. TASK NUMBER	
					5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way, Building 640 WPAFB OH 45433-8865					8. PERFORMING ORGANIZATION REPORT NUMBER AFIT/ISE/ENV/12-J04	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Intentionally Left Blank					10. SPONSOR/MONITOR'S ACRONYM(S)	
					11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT DISTRIBUTION STATEMENT A. APPROVED FOR PUBLIC RELEASE; DISTRIBUTION IS UNLIMITED.						
13. SUPPLEMENTARY NOTES This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States.						
14. ABSTRACT This research seeks to identify sources of conflict between operational realism requirements and analytical rigor requirements in defense Modeling and Simulation (M&S) efforts, and to provide recommendations which help alleviate identified conflicts. This research focuses on methods that can be used to improve the development and implementation of operator in the loop (OITL) virtual environments intended for use in acquisition decision making or the evaluation of operational plans. It is believed that the reduction of conflict between operators and analysts will lead to a better use of scarce M&S resources and produce better analytic results from M&S studies used as a basis for defense acquisition decision making. A real-world defense acquisition M&S case study is provided as an illustrative example from which recommendations and lessons learned are derived.						
15. SUBJECT TERMS Operational Realism, Design of Experiments, Modeling and Simulation, Analytic Rigor						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 35	19a. NAME OF RESPONSIBLE PERSON Dr. David R. Jacques (ENV)	
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (Include area code) (937)255-3636 x3329 david.jacques@afit.edu	

